Bias-Variance Tradeoff & Model Complexity

Nipun Batra and teaching staff

IIT Gandhinagar

July 30, 2025

Outline

Decision Tree Depth: A Story

Observation from Decision Trees:

- As depth increases, training accuracy **improves**

Decision Tree Depth: A Story

Observation from Decision Trees:

- As depth increases, training accuracy **improves**

Decision Tree Depth: A Story

Observation from Decision Trees:

- As depth increases, training accuracy **improves**
- As depth increases, test accuracy **improves up to a point**

Decision Tree Depth: A Story

Observation from Decision Trees:

- As depth increases, training accuracy **improves**
- As depth increases, test accuracy **improves up to a point**

Decision Tree Depth: A Story

Observation from Decision Trees:

- As depth increases, training accuracy **improves**
- As depth increases, test accuracy **improves up to a point**
- At very high depths, test accuracy **degrades** (overfitting!)

Decision Tree Depth: A Story

Observation from Decision Trees:

- As depth increases, training accuracy **improves**
- As depth increases, test accuracy **improves up to a point**
- At very high depths, test accuracy **degrades** (overfitting!)

Decision Tree Depth: A Story

Observation from Decision Trees:

- As depth increases, training accuracy **improves**
- As depth increases, test accuracy **improves up to a point**
- At very high depths, test accuracy **degrades** (overfitting!)

The Central Question

Why does this happen? How do we choose the right model complexity?

Pop Quiz: Overfitting Intuition

Quick Quiz 1

A decision tree with 1000 levels perfectly classifies all training data but performs poorly on test data. This is likely due to:

a) Underfitting - the model is too simple

Answer: b) The model memorized training specifics instead of learning general patterns!

Pop Quiz: Overfitting Intuition

Quick Quiz 1

A decision tree with 1000 levels perfectly classifies all training data but performs poorly on test data. This is likely due to:

a) Underfitting - the model is too simple

b) Overfitting - the model memorized training noise

Answer: b) The model memorized training specifics instead of learning general patterns!

Pop Quiz: Overfitting Intuition

Quick Quiz 1

A decision tree with 1000 levels perfectly classifies all training data but performs poorly on test data. This is likely due to:

a) Underfitting - the model is too simple
b) Overfitting - the model memorized training noise
c) Perfect generalization

Answer: b) The model memorized training specifics instead of learning general patterns!

The Fundamental Decomposition

For any learning algorithm, the expected test error can be decomposed as:

Bias-Variance Decomposition

$$\text{Expected Error} = \text{Bias}^2 + \text{Variance} + \text{Noise}$$

The Fundamental Decomposition

For any learning algorithm, the expected test error can be decomposed as:

Bias-Variance Decomposition

$$\text{Expected Error} = \text{Bias}^2 + \text{Variance} + \text{Noise}$$

The Fundamental Decomposition

For any learning algorithm, the expected test error can be decomposed as:

Bias-Variance Decomposition

$$\text{Expected Error} = \text{Bias}^2 + \text{Variance} + \text{Noise}$$

Component Definitions:

- **Bias**: Error due to overly simplistic assumptions

The Fundamental Decomposition

For any learning algorithm, the expected test error can be decomposed as:

Bias-Variance Decomposition

$$\text{Expected Error} = \text{Bias}^2 + \text{Variance} + \text{Noise}$$

Component Definitions:

- **Bias**: Error due to overly simplistic assumptions
- **Variance**: Error due to sensitivity to small changes in training set

The Fundamental Decomposition

For any learning algorithm, the expected test error can be decomposed as:

Bias-Variance Decomposition

$$\text{Expected Error} = \text{Bias}^2 + \text{Variance} + \text{Noise}$$

Component Definitions:

- **Bias**: Error due to overly simplistic assumptions
- **Variance**: Error due to sensitivity to small changes in training set
- **Noise**: Irreducible error in the data itself

Bias vs Variance: Visual Intuition

Bias vs Variance: Visual Intuition

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data
- Low complexity models

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data
- Low complexity models

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data
- Low complexity models

High Variance (Overfitting)

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data
- Low complexity models

High Variance (Overfitting)

- Model too complex

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data
- Low complexity models

High Variance (Overfitting)

- Model too complex
- Predictions vary wildly with training data

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data
- Low complexity models

High Variance (Overfitting)

- Model too complex
- Predictions vary wildly with training data
- **Example:** Deep decision tree

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data
- Low complexity models

High Variance (Overfitting)

- Model too complex
- Predictions vary wildly with training data
- **Example:** Deep decision tree
- High complexity models

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data
- Low complexity models

High Variance (Overfitting)

- Model too complex
- Predictions vary wildly with training data
- **Example:** Deep decision tree
- High complexity models

Bias vs Variance: Visual Intuition

High Bias (Underfitting)

- Model too simple
- Consistently wrong predictions
- **Example:** Linear model for complex non-linear data
- Low complexity models

High Variance (Overfitting)

- Model too complex
- Predictions vary wildly with training data
- **Example:** Deep decision tree
- High complexity models

The Tradeoff

Key insight: Reducing bias typically increases variance, and vice versa!

Pop Quiz: Bias-Variance Examples

Quick Quiz 2

Which scenario represents high variance?

a) A linear model consistently predicts poorly on both training and test sets

Answer: b) High variance means the model is too sensitive to training data changes!

Pop Quiz: Bias-Variance Examples

Quick Quiz 2

Which scenario represents high variance?

a) A linear model consistently predicts poorly on both training and test sets
b) A model gives very different predictions when trained on slightly different datasets

Answer: b) High variance means the model is too sensitive to training data changes!

Pop Quiz: Bias-Variance Examples

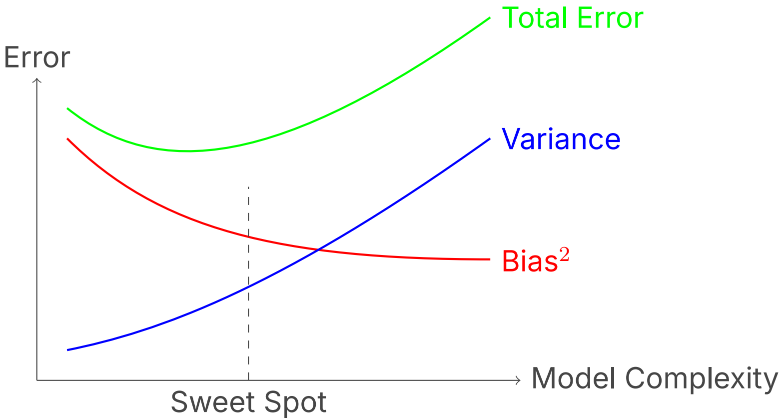
Quick Quiz 2

Which scenario represents high variance?

a) A linear model consistently predicts poorly on both training and test sets
b) A model gives very different predictions when trained on slightly different datasets
c) A model that always predicts the average target value

Answer: b) High variance means the model is too sensitive to training data changes!

The Model Complexity Spectrum



Goal: Find the complexity that minimizes total error!

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

Why We Need Cross-Validation

The Problem: We can’t directly optimize test error (we don’t have access to test set during training!)
The Solution: Cross-validation

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds
3. Validate on remaining fold

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds
3. Validate on remaining fold
4. Repeat k times

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds
3. Validate on remaining fold
4. Repeat k times
5. Average validation errors

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds
3. Validate on remaining fold
4. Repeat k times
5. Average validation errors

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds
3. Validate on remaining fold
4. Repeat k times
5. Average validation errors

Benefits

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds
3. Validate on remaining fold
4. Repeat k times
5. Average validation errors

Benefits

- Better error estimation

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds
3. Validate on remaining fold
4. Repeat k times
5. Average validation errors

Benefits

- Better error estimation
- Use all data for both training and validation

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds
3. Validate on remaining fold
4. Repeat k times
5. Average validation errors

Benefits

- Better error estimation
- Use all data for both training and validation
- Reduces impact of random splits

Why We Need Cross-Validation

The Problem: We can't directly optimize test error (we don't have access to test set during training!)
The Solution: Cross-validation

k-Fold Cross-Validation

1. Split data into k folds
2. Train on k-1 folds
3. Validate on remaining fold
4. Repeat k times
5. Average validation errors

Benefits

- Better error estimation
- Use all data for both training and validation
- Reduces impact of random splits
- Essential for model selection

Pop Quiz: Cross-Validation

Quick Quiz 3

In 5-fold cross-validation with 1000 data points, how many points are used for training in each fold?

a) 200 points

Answer: b) 800 points (4 out of 5 folds are used for training each time)

Pop Quiz: Cross-Validation

Quick Quiz 3

In 5-fold cross-validation with 1000 data points, how many points are used for training in each fold?

a) 200 points
b) 800 points

Answer: b) 800 points (4 out of 5 folds are used for training each time)

Pop Quiz: Cross-Validation

Quick Quiz 3

In 5-fold cross-validation with 1000 data points, how many points are used for training in each fold?

a) 200 points
b) 800 points
c) 1000 points

Answer: b) 800 points (4 out of 5 folds are used for training each time)

Model Selection Guidelines

- **High Bias (Underfitting):**

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high
- **High Variance (Overfitting):**

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high
- **High Variance (Overfitting):**
  - Decrease model complexity

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high
- **High Variance (Overfitting):**
  - Decrease model complexity
  - Use regularization

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high
- **High Variance (Overfitting):**
  - Decrease model complexity
  - Use regularization
  - Get more training data

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high
- **High Variance (Overfitting):**
  - Decrease model complexity
  - Use regularization
  - Get more training data
  - Large gap between training and validation error

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high
- **High Variance (Overfitting):**
  - Decrease model complexity
  - Use regularization
  - Get more training data
  - Large gap between training and validation error

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high
- **High Variance (Overfitting):**
  - Decrease model complexity
  - Use regularization
  - Get more training data
  - Large gap between training and validation error
- **Good Fit:**

Model Selection Guidelines

- **High Bias (Underfitting):**
  - Increase model complexity
  - Add more features
  - Use more sophisticated algorithms
  - Training and validation error both high
- **High Variance (Overfitting):**
  - Decrease model complexity
  - Use regularization
  - Get more training data
  - Large gap between training and validation error
- **Good Fit:**
  - Training and validation errors are close and reasonably low

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high
- Validation error: high

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high
- Validation error: high
- Small gap between them

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high
- Validation error: high
- Small gap between them
- More data doesn't help much

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high
- Validation error: high
- Small gap between them
- More data doesn't help much

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high
- Validation error: high
- Small gap between them
- More data doesn't help much

High Variance

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high
- Validation error: high
- Small gap between them
- More data doesn't help much

High Variance

- Training error: low

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high
- Validation error: high
- Small gap between them
- More data doesn't help much

High Variance

- Training error: low
- Validation error: high

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high
- Validation error: high
- Small gap between them
- More data doesn't help much

High Variance

- Training error: low
- Validation error: high
- Large gap between them

Debugging Learning Algorithms

Learning curves help diagnose bias vs variance:

High Bias

- Training error: high
- Validation error: high
- Small gap between them
- More data doesn't help much

High Variance

- Training error: low
- Validation error: high
- Large gap between them
- More data can help

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance
- **Cross-Validation:**

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance
- **Cross-Validation:**
  - Essential for model selection

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance
- **Cross-Validation:**
  - Essential for model selection
  - Provides reliable error estimates

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance
- **Cross-Validation:**
  - Essential for model selection
  - Provides reliable error estimates
  - Helps identify overfitting/underfitting

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance
- **Cross-Validation:**
  - Essential for model selection
  - Provides reliable error estimates
  - Helps identify overfitting/underfitting

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance
- **Cross-Validation:**
  - Essential for model selection
  - Provides reliable error estimates
  - Helps identify overfitting/underfitting
- **Practical Strategy:**

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance
- **Cross-Validation:**
  - Essential for model selection
  - Provides reliable error estimates
  - Helps identify overfitting/underfitting
- **Practical Strategy:**
  - Use learning curves to diagnose problems

Summary: Bias-Variance Tradeoff

- **Fundamental Tradeoff:**
  - Cannot reduce both bias and variance simultaneously
  - Must find optimal balance for your problem
- **Model Complexity:**
  - Simple models: High bias, low variance
  - Complex models: Low bias, high variance
- **Cross-Validation:**
  - Essential for model selection
  - Provides reliable error estimates
  - Helps identify overfitting/underfitting
- **Practical Strategy:**
  - Use learning curves to diagnose problems
  - Apply appropriate remedies based on bias vs variance

Next Steps: Ensemble Methods

How can we potentially reduce both bias and variance?

- **Ensemble Methods:** Combine multiple models

Next Steps: Ensemble Methods

How can we potentially reduce both bias and variance?

- **Ensemble Methods:** Combine multiple models

Next Steps: Ensemble Methods

How can we potentially reduce both bias and variance?

- **Ensemble Methods:** Combine multiple models
- **Reduce Variance:** Bagging, Random Forests

Next Steps: Ensemble Methods

How can we potentially reduce both bias and variance?

- **Ensemble Methods:** Combine multiple models
- **Reduce Variance:** Bagging, Random Forests

Next Steps: Ensemble Methods

How can we potentially reduce both bias and variance?

- **Ensemble Methods:** Combine multiple models
- **Reduce Variance:** Bagging, Random Forests
- **Reduce Bias:** Boosting, Gradient Boosting

Next Steps: Ensemble Methods

How can we potentially reduce both bias and variance?

- **Ensemble Methods:** Combine multiple models
- **Reduce Variance:** Bagging, Random Forests
- **Reduce Bias:** Boosting, Gradient Boosting

Next Steps: Ensemble Methods

How can we potentially reduce both bias and variance?

- **Ensemble Methods:** Combine multiple models
- **Reduce Variance:** Bagging, Random Forests
- **Reduce Bias:** Boosting, Gradient Boosting
- **Best of Both:** Modern ensemble techniques

Next Steps: Ensemble Methods

How can we potentially reduce both bias and variance?

- **Ensemble Methods:** Combine multiple models
- **Reduce Variance:** Bagging, Random Forests
- **Reduce Bias:** Boosting, Gradient Boosting
- **Best of Both:** Modern ensemble techniques

Next Steps: Ensemble Methods

How can we potentially reduce both bias and variance?

- **Ensemble Methods:** Combine multiple models
- **Reduce Variance:** Bagging, Random Forests
- **Reduce Bias:** Boosting, Gradient Boosting
- **Best of Both:** Modern ensemble techniques

Coming Up

Learn how ensemble methods can break the traditional bias-variance tradeoff!